A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP3145.

Exposures Related to House Dust Microbiota in a U.S. Farming Population

Mi Kyeong Lee,[1] Megan U. Carnes,[1] Natasha Butz,[2] M. Andrea Azcarate-Peril,[2] Marie Richards,[3] David M. Umbach,[4] Peter S. Thorne,[5] Laura E. Beane Freeman,[6] Shyamal D. Peddada,[4] and Stephanie J. London[1]

[1]Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Dept. of Health and Human Services, Research Triangle Park, North Carolina, USA
[2]Dept. of Medicine and Microbiome Core Facility, Center for Gastrointestinal Biology and Disease, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[3]Westat, Durham, North Carolina, USA
[4]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Dept. of Health and Human Services, Research Triangle Park, North Carolina, USA
[5]Department of Occupational and Environmental Health, University of Iowa, Iowa City, Iowa, USA
[6]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Dept. of Health and Human Services, Rockville, Maryland, USA

Background: Environmental factors can influence the house dust microbiota, which may impact health outcomes. Little is known about how farming exposures impact the indoor microbiota.

Objective: We aimed to identify exposures related to bacterial communities in house dust in a U.S. farming population.

Methods: We used 16S rRNA amplicon sequencing to characterize bacterial communities in vacuumed dust samples from the bedrooms of a subset of 879 households of farmers and farmers' spouses enrolled in the Agricultural Lung Health Study (ALHS), a case–control study of asthma nested within the Agricultural Health Study (AHS) in North Carolina and Iowa. Information on current farming (past 12 mo), including both crop and animal farming, and other potential microbial sources was obtained via questionnaires. We used linear regression to evaluate associations between exposures and bacterial diversity within each sample, analysis of similarity (ANOSIM), and permutational multivariate analysis of variance (PERMANOVA) to identify exposures related to diversity between samples, and analysis of composition of microbiome to examine whether exposures related to diversity were also related to differential abundance of specific operational taxonomic units (OTUs).

Results: Current farming was positively associated with bacterial diversity in house dust, with or without adjustment for nonfarm exposures related to diversity, including presence of indoor pets, home condition, and season of dust collection. Many taxa exhibited differential abundance related to farming. Some taxa in the phyla Chloroflexi and Verrucomicrobia were associated [false discovery rate (FDR) < 0.05] with farming but not with other nonfarm factors. Many taxa correlated with the concentration of house dust of endotoxin, commonly studied as a general marker of exposure to the farming environment.

Conclusions: In this farming population, house dust microbiota differed by current farming status. Understanding the determinants of the indoor microbiota is the first step toward understanding potential relationships with health outcomes. https://doi.org/10.1289/EHP3145

Introduction

House dust contains a wide range of microorganisms. The house dust microbiota is influenced by environmental factors and, in turn, can impact human health (Ownby et al. 2002; von Mutius and Vercelli 2010). Understanding the determinants of the indoor dust microbiota is necessary before examining potential associations with health outcomes. In general populations, factors known to be associated with bacterial communities include the presence of indoor pets and the number of occupants (Barberán et al. 2015; Dannemiller et al. 2016). Although farming populations have unique microbial exposures, few studies have explored the determinants of the microbiota in these settings. In a recent study of 86 children, farm exposures were associated with bacterial communities of dust inside homes (Birzele et al. 2017). There have been no large-scale studies exploring determinants of the house dust microbiota in a farming population.

Endotoxin is a proxy measure of bacteria exposure that consists of lipopolysaccharide found in the outer membrane of Gram-negative bacteria. Studies have measured endotoxin concentrations in house dust and identified associations with various health endpoints (Liebers et al. 2008). Studies in farming populations, conducted mostly in European countries, have reported that the presence of pets, living on a farm, cleanliness, geographical region, and season (Chen et al. 2012; Dassonville et al. 2008; Giovannangelo et al. 2007; Holst et al. 2015) are associated with endotoxin concentration in house dust. However, no studies have examined how house dust endotoxin relates to specific bacterial taxa.

We hypothesized that current farming, defined as working with crops or animals during the past 12 mo, influences the microbiota inside homes. We examined associations between environmental exposures and bacterial communities in bedroom dust from 879 independent households in the Agricultural Lung Health Study (ALHS), a study of farmers and farmers' spouses in North Carolina and Iowa. We evaluated bacterial diversity in association with farm exposures (such as living on a farm, current farming, and types of farming) and nonfarm exposures (such as demographic characteristics, presence of indoor pets, and home condition) among the study population. For exposures associated with bacterial diversity, we identified specific operational taxonomic units (OTUs) that showed differential abundance according to these exposures. In addition, we examined associations of house dust endotoxin with bacterial diversity and with relative abundance of OTUs.

Methods

Study Population

The ALHS is a case–control study of adult asthma nested within the Agricultural Health Study (AHS), a prospective cohort of licensed

Address correspondence to S.J. London, National Institute of Environmental Health Sciences, P.O. Box 12233, Mail Drop A3-05, Research Triangle Park, NC 27709, USA. Telephone: (984) 287-3688. Email: london2@niehs.nih.gov

Supplemental Material is available online (https://doi.org/10.1289/EHP3145).

The authors declare they have no actual or potential competing financial interests.

Received 22 November 2017; Revised 1 May 2018; Accepted 2 May 2018; Published 1 June 2018.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

pesticide applicators ($n = 52,394$ private applicators), mostly farmers and their spouses ($n = 32,345$) enrolled from 1993 to 1997 in North Carolina and Iowa (Alavanja et al. 1996). The farmers in the study are individuals who reported their occupation as a farmer on the initial enrollment questionnaire. Details of ALHS subject selection have been previously described (Carnes et al. 2017; House et al. 2017). In brief, potential ALHS participants were identified from among the 44,130 respondents (24,171 farmers and 19,959 spouses) to the AHS follow-up questionnaire administered by phone from 2005 to 2010 (data version P3REL201209.00) based on responses to questions about asthma diagnosis and symptoms. A total of 3,301 participants were enrolled in the ALHS (1,223 with asthma, response rate = 51.7% and 2,078 noncases, response rate = 50.0%). As previously described (House et al. 2017), we used three definitions of asthma to avoid missing undiagnosed asthmatics. Most enrolled cases ($n = 876$) responded "yes" to the two questions: "Have you ever been diagnosed with asthma?" and "Do you still have asthma?" and "no" to the following two questions: "Have you ever been diagnosed with chronic obstructive pulmonary disease (COPD)?" and "Have you ever been diagnosed with emphysema?" We also identified never smoking ($n = 263$) or minimal past smoking (≤ 10 pack-years, $n = 46$) cases of likely undiagnosed asthma (total = 309) based on report of current asthma symptoms and use of asthma medications and "no" to the above questions regarding diagnosis of either COPD or emphysema. Further, because asthma and COPD can coexist, we also enrolled 38 subjects reporting current asthma and prior diagnosis of either COPD or emphysema as long as they were never-smokers ($n = 28$) or past-smokers (≤ 10 pack-years, $n = 10$). Noncases were randomly chosen from among individuals not categorized as cases.

Of the 3,301 participants, 2,871 received a home visit at which bedroom dust was collected. For 2,692 of these, dust was sufficient for endotoxin measurement. We sent a random sample ($n = 1,000$ out of 2,692, including 372 asthma cases) for measurement of house dust microbiota. Workflow of the study including selection criteria, and number of samples can be found in Figure 1.

The study was approved by the Institutional Review Board at the National Institute of Environmental Health Sciences. Written informed consent was obtained from all participants.

A dust sample was collected from the bedroom by a trained field technician during the home visit as previously described (Carnes et al. 2017). We used a DUSTREAM® Collector (Indoor Biotechnologies, Inc.) to vacuum a 1-yd^2 (0.84-m^2) area on the sleeping surface and on the floor next to the bed for 2 min in each area. Dust samples were sent to Social & Scientific Systems, Inc. to be sieved, weighted into aliquots of 50 mg, and frozen at −20°C.

Dust samples were extracted and endotoxin was measured using the *Limulus* amebocyte lysate assay (Lonza Walkersville, Inc.) at the University of Iowa (Thorne et al. 2005; Thorne et al. 2015; Vojta et al. 2002). Reagents from a single lot were used in all assays, each of which also included low- and high-endotoxin quality control house dust samples. Values below the limit of detection (LOD) [0.00048 endotoxin units (EU)/mg, $n = 2$ of 879] were assigned to LOD divided by the square root of two (0.00034) (Carnes et al. 2017).

DNA Isolation and 16S rRNA Amplicon Sequencing

Samples were randomized before processing. DNA was isolated using the Mo Bio 96 well plate PowerSoil DNA extraction kit (QIAGEN, Inc.). The DNA isolation protocol was applied as recommended by the manufacturer with the following modifications: approximately 0.3–0.5 g of each sample were loaded into each well of the PowerSoil bead beating plate. Samples were incubated

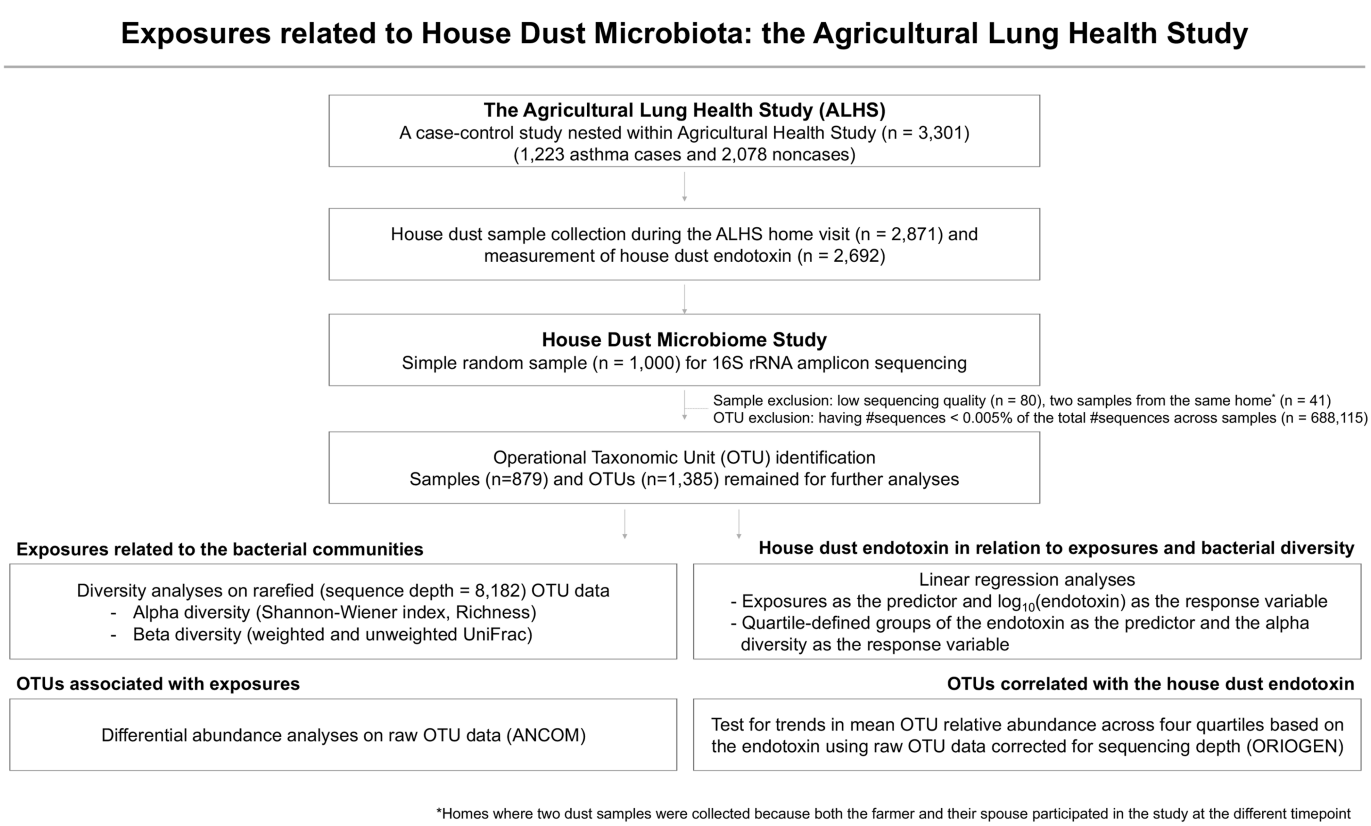

Figure 1. Workflow of our house dust microbiome study. This workflow includes a summary of sample selection from the Agricultural Lung Health Study (ALHS) ($n = 3,301$) to the house dust microbiome study ($n = 879$). It shows association analyses used in this paper: bacterial diversity analysis for both environmental exposures and endotoxin concentration, differential abundance analysis for environmental exposures, and differential relative abundance analysis for endotoxin.

in PowerSoil bead solution and buffer C1 at 70°C for 20 min before bead beating to aid in lysis of spores, and 50 μL of elution buffer C6 was used. Then, post-DNA extraction, each sample was quantified using the NanoDrop™ (A260) (Thermo Fisher Scientific Inc.) and normalized to 5 ng/μL DNA per sample before PCR amplification of V3–V4 region of the *16S rRNA* gene.

DNA was amplified using primers targeting the V3–V4 region of the bacterial *16S rRNA* gene and overhang adapter sequences appended to the primer pair for compatibility with Illumina index and sequencing adapters. The complete sequences of the primers were: F–5′ GTGCCAGCAGCCGCGGTAA-3′ and R–5′ GTCT CGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTAC HVGGGTWTCTAAT-3′. Master mixes contained 12.5 ng of total DNA and 2X KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Inc.). The thermal profile for the amplification of each sample had an initial denaturing step at 90°C for 3 min, followed by a cycling of denaturing of 95°C for 30 s, annealing at 55°C for 30 s and a 30-s extension at 72°C (25 cycles), a 5-min extension at 72°C, and a final hold at 4°C. Each 16S amplicon was purified using AMPure XP reagent (Beckman Coulter, Inc.). In the next step, each sample was amplified using a limited cycle PCR program, adding Illumina sequencing adapters and dual-index barcodes [index 1(i7) and index 2(i5)] (Illumina) to the amplicon target. For the second round of amplification, the thermal profile consisted of an initial denaturing step at 95°C for 3 min, followed by a denaturing cycle of 95°C for 30 s, annealing at 55°C for 30 s and a 30 s extension at 72°C (8 cycles), and 5-min extension at 72°C. The final libraries were again purified using AMPure XP reagent (Beckman Coulter, Inc.), quantified, and normalized before pooling. The DNA library pool was then denatured with NaOH, diluted with hybridization buffer, and heat denatured before loading on the MiSeq reagent cartridge (Illumina) and on the MiSeq instrument (Illumina). Automated cluster generation and paired-end sequencing with dual reads were performed per manufacturer's instructions.

Preprocessing of sequencing reads included paired-end joining, demultiplexing, and quality filtering using Illumina software and fastq-join. Paired-end fastq files were generated from the sequencing results of an Illumina MiSeq using configureBclToFastq. The paired-end fastqs were joined into a single multiplexed, single-end fastq using the software tool fastq-join. Demultiplexing and quality filtering were performed on the joined results. Quality control reports were produced using the software FastQC (FastQC version 0.11.2). Measurements were made by the Microbiome Core Facility at the University of North Carolina.

To generate bacterial community information from filtered sequencing data, we used Quantitative Insights Into Microbial Ecology (QIIME; version 1.9.1) (Caporaso et al. 2010b). *De novo* OTU picking was performed using the script pick_de_novo_otus.py with UCLUST (Edgar 2010), the default OTU clustering algorithm. To remove chimeric sequences, we used identify_chimeric_seqs.py to run ChimeraSlayer (Haas et al. 2011) after aligning sequences using align_seqs.py with PyNAST (Caporaso et al. 2010a), the default alignment method. Assign_taxonomy.py with the Greengenes database (version 13_5; greengenes.secondgenome.com) was used to assign taxonomy to each sequence. A phylogenetic tree was constructed using make_phylogeny.py with FastTree (Price et al. 2009); the default phylogeny construction algorithm. OTUs were assigned into seven taxonomic levels: kingdom, phylum, class, order, family, genus, and species.

Quality filtering criteria for the OTU data included (*a*) exclusion of samples having <10,000 sequencing depth (Barberán et al. 2015), and (*b*) removal of chimeric sequences and OTUs having <0.005% of the total number of sequence reads (Bokulich et al. 2013; Navas-Molina et al. 2013). During the filtering, 80 samples and 604,569 OTUs were excluded after removal of 83,546 OTUs of chimeric sequences. For 41 homes where two dust samples were collected because both the farmer and their spouse participated in the study at different time points, we chose the record from the farmer for further analyses so that all samples were independent. These exclusions left 879 samples (including 333 asthma cases and 546 noncases) for statistical analysis.

Environmental and Other Factors

We examined demographic, home condition, and environmental factors assessed at the home visit in relation to the indoor dust microbiota. Demographic factors included state of residence (North Carolina vs. Iowa) and gender. We assessed the presence of indoor pets in the past 12 mo (dogs or cats) by participant questionnaires. The home condition was rated by field technicians at the time of the visit using a five-point scale used in an earlier home allergen study (Arbes et al. 2003). The five categories were: (*a*) "Extremely Poor: lack of organization," (*b*) "Poor," (*c*) "Average: clean with moderate clutter," (*d*) "Above Average," and (*e*) "Good: organized, clean all over." We dichotomized responses into lower (levels 1–2) or higher home condition (levels 3–5). Because not all farmers or spouses of farmers are currently farming, we used questionnaire data to classify them with regard to current farm exposures. Subjects who responded "yes" to "Do you currently live on a farm?" were classified as living on a farm. Individuals who responded "yes" to "In the past year, have you worked with soybeans or grains?" were classified as currently crop farming. Individuals who responded "yes" to "In the past year, have you worked with any farm animals?" were classified as currently animal farming. Individuals who responded "yes" to this question were further asked, "In the past 12 mo, have you worked with (a) dairy cattle, (b) beef cattle, (c) hogs, (d) chicken or turkeys, and (e) other farm animals?" We defined current farming by either crop or animal farming. We combined information on current crop and animal farming to create a four-category variable: neither crop nor animal farming (reference category), crop farming only, animal farming only, or both crop and animal farming. Additionally, we calculated the number of types of farm animals (beef or dairy cattle, hogs, or poultry) that individuals worked with, and categorized it into three groups: 0 type, 1 type, and 2 + types. Season of dust collection (spring, summer, fall, winter) was generated based on the date of the home visit: March 21–June 20 for spring, June 21–September 20 for summer, September 21–December 20 for fall, and December 21–March 20 for winter.

Given that the study participants were randomly sampled from a nested case–control study of asthma, we also considered whether asthma status was related to the indoor dust microbiota.

Given that one of potential factors reported to be associated with the endotoxin is carpeting, we also examined whether carpeting was associated with bacterial communities and endotoxin levels in the house dust microbiota. The carpeting variable was obtained via questionnaire completed by the technician who collected the dust sample, and dichotomized into two categories: carpeted surface vs. smooth floor (no carpeted or no carpet sampled group).

Association Analyses

In association analyses both for microbial communities and endotoxin, we considered farm exposures, such as living on a farm, crop farming, and animal farming, and nonfarm exposures, such as state of residence, gender of study participants, presence of indoor pets, and home condition, unless noted otherwise. Each exposure was treated as a binary variable, comparing exposed to unexposed (for example: presence of indoor pets vs. no indoor pets, higher vs. lower home condition, crop farming vs. no crop farming, animal farming vs. no animal farming) unless stated otherwise. Regarding

season of dust collection, we compared one season vs. all other seasons combined (for example, spring vs. summer, fall, and winter combined; summer vs. spring, fall, and winter combined) unless stated otherwise.

State of residence was related to current farm exposures such as living on a farm, crop farming, and animal farming. Instead of adjusting for state, we performed a stratified analysis.

Bacterial diversity analysis using rarefied operational taxonomic unit data. For bacterial diversity within each sample (alpha diversity), we calculated the Shannon index (Shannon 1948), which reflects both the number of species and the distribution of the relative abundance of each species. We used linear regression to evaluate associations between exposures and this diversity measure. We also evaluated associations between farm exposures and diversity after adjusting for nonfarm exposures significantly associated with diversity in univariate analyses (p-value <0.05) by adding the nonfarm exposures as covariates in linear regression models. To assess associations of number of types of farm animals with diversity, we evaluated a linear trend between diversity and the ordered categories of 0, 1, or 2+ types by assigning scores of 1, 2, or 3 to those categories. The t-test associated with the linear term provided the test for trend.

Additionally, we have examined associations of bacterial (alpha) diversity with other factors that can be factors of interest to be considered in our study: asthma status and carpeting. As asthma status was the selection factor in the ALHS, we tested associations between asthma and bacterial communities with the Shannon index in linear regression with asthma status as the predictor and the bacterial diversity as the response variable. We also examined whether adjusting for asthma status with the other covariates changed the results. For carpeting with respect to associations with the bacterial diversity, we used linear regression with the carpeting variable as the predictor and the diversity index as the response variable.

To test whether bacterial community compositions between samples (beta diversity) differed across exposure levels, we used two different methods. One method was analysis of similarity (ANOSIM) (Clarke 1993) with the Bray-Curtis (BC) dissimilarity (Bray and Curtis 1957) metric; ANOSIM generates a statistic (denoted R) that ranges from 0 (no separation) to 1 (complete separation). The second method was permutational multivariate analysis of variance (PERMANOVA) (Anderson 2001) with the unique fraction (UniFrac) metric (Lozupone and Knight 2005), a measure of the phylogenetic distance among OTUs in a phylogenetic tree. PERMANOVA generates a statistic-denoted R^2 explanatory power, which estimates how much variability can be explained by an exposure. We used both weighted, taking abundances of OTUs into account, and unweighted UniFrac. The number of permutations was 999 for both ANOSIM and PERMANOVA.

To avoid any bias due to different sequencing depth among samples, the data were rarefied to the minimum number of sequences ($n=8{,}182$) across samples for diversity analyses. We used R (version 3.2.4; R Development Core Team) to summarize characteristics of the study population and perform the diversity analyses. We used functions including diversity, vegdist, anosim, and adonis in R package vegan (version 2.4.3; R Development Core Team) for computing the Shannon index, the BC dissimilarity metric, ANOSIM, and PERMANOVA, respectively. We used UniFracs in R package phyloseq (version 1.19.1; R Development Core Team). (McMurdie and Holmes 2013) for the UniFrac phylogenetic distance metric. We set a threshold of p-value <0.05 for statistical significance for the diversity analyses.

Inference about abundance using un-rarefied operational taxonomic unit data. To identify OTUs whose abundances differ significantly by exposure levels, we applied the analysis of composition of microbiomes (ANCOM) (Mandal et al. 2015) to unrarefied data because ANCOM's log-ratio approach accounts for variation in sequencing depth across samples. For both nonfarm and farm exposures, exposure was the predictor and the abundances the response variable. To identify OTUs significantly associated with current farming after adjusting for the nonfarm factors significantly associated with bacterial communities in the diversity analysis, we included the nonfarm factors as covariates in the models. We also examined differential abundance according to the four-level current farming (neither crop nor animal farming, crop farming only, animal farming only, and both crop and animal farming) by using the four-level variable as the predictor and the abundance as the response variable in ANCOM. To correct for multiple testing, we set the false discovery rate (FDR) to 0.05. We summarized the number of OTUs with FDR <0.05 at the phylum level. When a family includes at least two OTUs whose abundances significantly uniquely differed by either crop or animal farming, we calculated relative abundances at the family level.

Endotoxin in relation to environmental factors and bacterial diversity using rarified data. We examined the relationship of house dust endotoxin with both farm and nonfarm exposures using linear regression with exposure as the predictor and $\log_{10}$-transformed endotoxin levels as the response variable. To evaluate associations between farm exposures and the endotoxin levels after adjusting for nonfarm exposures significantly associated with the endotoxin in univariate analyses (p-value <0.05), we added the nonfarm exposures as covariates in linear regression models. In this analysis of environmental factors in relation to endotoxin levels, we also tested whether carpeting status (carpeted surface vs. smooth floor) was associated with the endotoxin levels by using linear regression. We used the carpeting variable as the predictor and the endotoxin levels as the response variable.

To evaluate associations between bacterial diversity (the Shannon index) and endotoxin, we created ordered categories based on endotoxin quartiles. We used linear regression with diversity as the response variable and with indicator variables for each endotoxin category as predictors (lowest quartile as reference). We tested for linear trend using a score variable whose value was 1, 2, 3, or 4 for endotoxin levels in the first through fourth quartile categories. The t-test associated with the linear term provided the test for trend.

Endotoxin in relation to relative abundance using unrarefied operational taxonomic unit data. We examined positive or negative trends in OTU-specific relative abundance across endotoxin quartiles. For each OTU, after correcting for the differential sequencing depth across samples (see below for details), we tested whether its relative abundance significantly increased (mean relative abundance in quartile 1 $\leq$ mean relative abundance in quartile 2 $\leq$ mean relative abundance in quartile 3 $\leq$ mean relative abundance in quartile 4) or decreased (mean relative abundance in quartile 1 $\geq$ mean relative abundance in quartile 2 $\geq$ mean relative abundance in quartile 3 $\geq$ mean relative abundance in quartile 4) over endotoxin quartiles. We used the software ORIOGEN (order-restricted inference for ordered gene expression) version 4.01 at an FDR threshold of 0.05 for performing these analyses (Peddada et al. 2005).

OTU counts were normalized to account for variations in sequencing depth. For this OTU-level analysis, raw OTU counts were transformed to account for within-sample and between-sample variations in sequencing depth. The normalized counts can be regarded as estimates of the relative abundance of the OTU in a sample; we refer to the normalized counts as relative abundance.

C_{ijk} denotes the raw count for the i^{th} OTU in the j^{th} sample of the k^{th} exposure group (e.g., endotoxin quartiles ranging from 1 to 4). We add 1 to each count to accommodate ln transformation of

possibly zero counts and take the average of the ln-transformed counts across all OTUs in a sample and exposure group. $T'_{\cdot jk}$ denotes that sample and exposure group–specific average of ln-transformed values, which can be expressed in symbols as:

$$T'_{\cdot jk} = \frac{1}{N}\sum_{i=1}^{N} \ln(C_{ijk}+1).$$

Thus, $T'_{\cdot jk}$ is the ln-transformed geometric mean of the $C_{ijk}+1$ for all OTUs in the sample and exposure group. We further average these averages across all the samples in an exposure group, namely,

$$T'_{\cdot\cdot k} = \frac{1}{n_k}\sum_{j=1}^{n_k} T'_{\cdot jk}.$$

Then we define the transformed count for OTU i in sample j and exposure group k as:

$$C'_{ijk} = \ln(C_{ijk}+1) - [\,T'_{\cdot jk} - T'_{\cdot\cdot k}].$$

The term in square brackets can be regarded as a ln-transformed adjusted geometric mean count for the sample and exposure group. The result of subtracting the ln-transformed adjusted geometric mean count from the ln-transformed raw count is a ln-transformed ratio: raw count divided by adjusted geometric mean count. In this sense, the transformed count C'_{ijk} can be interpreted as assessing a ln-transformed relative abundance for the i^{th} OTU in the j^{th} sample of the k^{th} exposure group.

Results

Association analyses included dust samples from 879 independent homes from North Carolina (32%) and Iowa (68%). Participants were either farmers (532, 97% male) or the spouse of a farmer (347) (Table 1). Of the 532 farmers, 171 (158 males and 13 females) were from North Carolina, and 361 (358 males and three females) were from Iowa. Of the 347 spouses of farmers, one was male, and the remaining were female spouses. Of the 532 farmers, 462 were married, and 2 were living as married. Participants were 62 y old on average at the home visit. Indoor pets (dogs or cats) were present in 42% of homes. Most homes (78%) were rated in the higher of the two home condition categories by the field technician. Participants reported living on a farm (83%), crop farming in the past 12 mo (55%), and animal farming in the past 12 mo (50%). For the combined crop or animal current farming variable, 35% of participants reported that they were not currently farming, 15% reported crop farming only, 10% reported animal farming only, and 40% reported both crop and animal farming. Of the 305 participants who were not currently farming (neither crop nor animal), 214 were female, and 91 were male. Of the 214, 205 were spouses of farmers, and 9 were farmers. The 91 males were all farmers. The geometric mean endotoxin level in house dust was 38.4 EU/mg (median, 51.0 EU/mg; minimum, 0.00034 EU/mg; 25th percentile, 24.5 EU/mg; 75th percentile, 84.4 EU/mg; and maximum 1,170 EU/mg). There were 333 current asthma cases in our study: 99 from North Carolina and 234 from Iowa. Of the 333, 178 (50 from North Carolina and 128 from Iowa) were farmers, and 155 (49 from North Carolina and 106 from Iowa) were spouses of farmers.

After quality filtering on the sequencing depth, the total number of sequences across all samples was 72,865,099. The number of sequences per sample ranged from 8,182 to 448,744 (mean = 82,895; median = 73,708). Of all OTUs across all samples, 73% had nonzero counts. Based on 97% sequence similarity, there were 1,385 OTUs for association analyses. Of these, at the taxonomic level of kingdom, 1,346 OTUs were assigned to Bacteria, seven to Archaea, and 32 were unassigned. The number of unassigned OTUs in each taxonomic level were 32 (2.3%) for phylum, 35 (2.5%) for class, 56 (4.0%) for order, 198 (14.3%) for family, 656 (47.4%) for genus, and 1,287 (92.9%) for species.

With respect to asthma status, the selection factor in the ALHS, we evaluated association of the disease status with bacterial communities. No association was observed between asthma and our measure of bacterial diversity, Shannon index: mean Shannon index = 4.58 for both cases and noncases, and p for difference = 0.99. Although asthma status was clearly not related to diversity, we additionally verified that adding asthma to the models with the other covariates did not change the results. For example, for living on a farm, $p = 0.525$ before and $p = 0.526$ after adjustment; for crop farming, $p = 0.006$ before and 0.006 after adjustment; and for animal farming, $p = 3.4 \times 10^{-5}$ before and $p = 3.5 \times 10^{-5}$ after adjustment. Therefore, we did not include asthma status in further analyses.

Several environmental factors were significantly associated with alpha diversity. Higher Shannon index (one metric of alpha diversity), reflecting both the number of species and their relative abundance, was associated with residence in Iowa (as opposed to North Carolina), presence of an indoor dog, crop and animal farming, and winter (vs. other seasons combined). Lower Shannon index was related to home condition rated in the higher category (Table 2). Significantly higher diversity was observed in homes of individuals working with beef cattle, dairy cattle, or poultry compared to those not working with these animals. Further, the number of types of farm animals (beef or dairy cattle, hogs, or poultry) that individuals worked with was positively related to Shannon index [mean index ± standard deviation (SD) for 0 type = 4.50 ± 0.76, for 1 type = 4.66 ± 0.72, and for 2+ types = 4.71 ± 0.73; p-value for linear trend = 0.003]. When we performed stratified analyses by state of residence (North Carolina or Iowa), these patterns were similar in both, but given the larger sample size in Iowa ($n = 596$), results for Iowa tended to be more highly statistically significant than those for North Carolina ($n = 283$) (Table S1). After we adjusted for presence of an indoor dog, home condition, and winter, the associations between alpha diversity measure (Shannon index) and current farming remained significant (Table S2).

For beta diversity (measure of between-sample difference), when we assessed differences between demographic or exposure groups based on BC dissimilarity, we observed differences for state of residence, presence of indoor pets, current crop farming, current animal farming, and season of dust collection (spring vs. other seasons combined) (Table S3). The measures of group separation (R statistics ranging from 0 for no separation to 1 for complete separation) were small but statistically significant (p-value <0.01). R values for the state of residence (0.128) and animal farming (0.055) were relatively greater than those for other factors: 0.016 for presence of indoor pets, 0.044 for season of dust collection (spring vs. other seasons combined), and 0.038 for current crop farming. When we assessed beta diversity by differences between groups based on phylogenetic distances (weighted UniFrac), microbial communities were significantly different by most factors: state of residence, gender, presence of indoor pets, home condition, current farming, and season of dust collection (Table S4). Of note, we observed relatively greater values of R^2 for state of residence ($R^2 = 0.028$ from weighted approach; $R^2 = 0.043$ from unweighted approach) and for animal farming ($R^2 = 0.012$ from weighted approach; $R^2 = 0.019$ for unweighted approach) compared to those for other variables ranging between 0.002 and 0.014 from either of the approaches. Of total variability of bacterial communities in house dust, 3% or 4% can be explained by the state of residence and 1% or 2% by the animal farming exposure. Largely, analyses with unweighted UniFrac showed similar results but with slightly greater R^2 values compared to those with UniFrac weighted by OTU count (Table S4).

Table 1. Characteristics of the study participants.

Characteristics, n (%)[a]	North Carolina ($n = 283$)	Iowa ($n = 596$)	Total ($n = 879$)
Farmer or spouse at the time of AHS enrollment			
Farmer	171 (60.4)	361 (60.6)	532 (60.5)
Spouse of farmer	112 (39.6)	235 (39.4)	347 (39.5)
Gender			
Male	159 (56.2)	358 (60.1)	517 (58.8)
Female	124 (43.8)	238 (39.9)	362 (41.2)
Age, years (mean ± SD)	63 ± 11	61 ± 11	62 ± 11
Endotoxin[b], EU/mg [geometric mean(SE)]	31.77 (0.29)	42.04 (0.14)	38.42 (0.14)
Presence of indoor pets – past 12 months			
Dogs or cats			
Yes	131 (46.3)	240 (40.3)	371 (42.2)
No	152 (53.7)	356 (59.7)	508 (57.8)
Dogs			
Yes	103 (36.4)	167 (28.0)	270 (30.7)
No	180 (63.6)	429 (72.0)	609 (69.3)
Cats			
Yes	55 (19.4)	127 (21.3)	182 (20.7)
No	228 (80.6)	469 (78.7)	697 (79.3)
Home condition[c]			
Higher	208 (74.0)	474 (79.5)	682 (77.8)
Lower	73 (26.0)	122 (20.5)	195 (22.2)
Carpeting[d]			
Carpeted surface	257 (91.1)	563 (94.6)	820 (93.5)
Smooth floor	25 (8.9)	32 (5.4)	57 (6.5)
Current[e] farm exposure			
Living on a farm			
Yes	220 (77.7)	507 (85.1)	727 (82.7)
No	63 (22.3)	89 (14.9)	152 (17.3)
Crop farming			
Yes	94 (33.2)	386 (64.8)	480 (54.6)
No	189 (66.8)	210 (35.2)	399 (45.4)
Animal farming			
Yes	113 (39.9)	330 (55.4)	443 (50.4)
No	170 (60.1)	266 (44.6)	436 (49.6)
Beef cattle			
Yes	76 (26.9)	233 (39.1)	309 (35.2)
No	207 (73.1)	363 (60.9)	570 (64.8)
Dairy cattle			
Yes	8 (2.8)	46 (7.7)	54 (6.1)
No	275 (97.2)	550 (92.3)	825 (93.9)
Hogs			
Yes	19 (6.7)	114 (19.1)	133 (15.1)
No	264 (93.3)	482 (80.9)	746 (84.9)
Poultry			
Yes	40 (14.1)	60 (10.1)	100 (11.4)
No	243 (85.9)	536 (89.9)	779 (88.6)
Current farming: crops and/or animals			
Neither crop nor animal farming	140 (49.5)	165 (27.7)	305 (34.7)
Crop farming only	30 (10.6)	101 (16.9)	131 (14.9)
Animal farming only	49 (17.3)	45 (7.6)	94 (10.7)
Both crop and animal farming	64 (22.6)	285 (47.8)	349 (39.7)
Number of types of farm animals			
0 type	179 (63.3)	284 (47.7)	463 (52.7)
1 type	71 (25.1)	202 (33.9)	273 (31.1)
2 + types	33 (11.7)	110 (18.5)	143 (16.3)
Season of dust collection			
Spring, March 21–June 20	78 (27.6)	145 (24.3)	223 (25.4)
Summer, June 21–September 20	74 (26.1)	191 (32)	265 (30.1)
Fall, September 21–December 20	56 (19.8)	133 (22.3)	189 (21.5)
Winter, December 21–March 20	75 (26.5)	127 (21.3)	202 (23.0)
Current asthma			
Cases	99 (35.0)	234 (39.3)	333 (37.9)
Noncases	184 (65.0)	362 (60.7)	546 (62.1)

Note: AHS, Agricultural Health Study; EU, endotoxin units; SD, standard deviation; SE, standard error.
[a]Percentages may not add to exactly 100 due to rounding.
[b]Endotoxin was measured using the *Limulus* amebocyte lysate assay.
[c]A field technician rated home condition at the time of the home visit using a five-point scale that we dichotomized into higher vs. lower. Two samples from North Carolina with missing home condition were removed for any analysis considering home condition.
[d]Two samples (one from North Carolina and one from Iowa) with missing carpeting status were removed for any analysis considering the carpeting. Smooth floor represents no carpeted ($n = 54$) or no carpet sampled ($n = 3$) group.
[e]Current defined as past 12 months. Some farmers are not currently doing farm work, and some spouses are currently doing farm work but are not farmers.

Table 2. Associations between alpha diversity (Shannon index) and exposures.

Environmental and other factors	Yes		No		
	n	Alpha diversity[a]	n	Alpha diversity	p[e]
State of residence (North Carolina = Yes; Iowa = No)	283	4.46 ± 0.79	596	4.64 ± 0.72	0.001
Gender (male = Yes; female = No)	517	4.56 ± 0.80	362	4.62 ± 0.68	0.219
Presence of indoor pets, past 12 months	—	—	—	—	—
Dogs or cats (vs. neither dogs nor cats)	371	4.63 ± 0.74	508	4.54 ± 0.75	0.081
Dogs (vs. no dogs)	270	4.67 ± 0.71	609	4.54 ± 0.76	0.020
Cats (vs. no cats)	182	4.62 ± 0.73	697	4.57 ± 0.76	0.410
Home condition, higher category[b] (vs. lower category)	682	4.54 ± 0.76	195	4.74 ± 0.70	5.8E-04
Carpeting, carpeted surface (vs. smooth floor)	820	4.57 ± 0.76	57	4.73 ± 0.65	0.134
Current[c] farm exposure	—	—	—	—	—
Living on a farm (vs. not living on a farm)	727	4.59 ± 0.76	152	4.54 ± 0.71	0.402
Crop farming (vs. no crop farming)	480	4.64 ± 0.74	399	4.51 ± 0.75	0.011
Animal farming (vs. no animal farming)	443	4.64 ± 0.71	436	4.47 ± 0.77	9.0E-06
Beef cattle (vs. no beef cattle)	309	4.67 ± 0.73	570	4.53 ± 0.76	0.009
Dairy cattle (vs. no dairy cattle)	54	4.90 ± 0.54	825	4.56 ± 0.76	0.001
Hogs (vs. no hogs)	133	4.67 ± 0.71	746	4.57 ± 0.75	0.137
Poultry (vs. no poultry)	100	4.73 ± 0.75	779	4.56 ± 0.75	0.038
Number of types of farm animals	—	—	—	—	—
0 type	—	4.50 ± 0.76		—	0.003[f]
1 type	—	4.66 ± 0.72	—	—	—
2+ types	—	4.71 ± 0.73	—	—	—
Season of dust collection[d]	—	—	—	—	—
Spring, March 21–June 20 (vs. other seasons combined)	223	4.54 ± 0.86	656	4.60 ± 0.71	0.327
Summer, June 21–September 20 (vs. other seasons combined)	265	4.53 ± 0.78	614	4.60 ± 0.73	0.201
Fall, September 21–December 20 (vs. other seasons combined)	189	4.52 ± 0.64	690	4.60 ± 0.78	0.224
Winter, December 21–March 20 (vs. other seasons combined)	202	4.75 ± 0.65	677	4.53 ± 0.77	3.1E-04
Current asthma (cases vs. noncases)	333	4.58 ± 0.08	546	4.58 ± 0.72	0.99

Note: Alpha diversity refers to bacterial diversity within each sample. We used rarefaction with the minimum sequencing depth across all samples (n = 8,182).
[a]Average ± standard deviation of Shannon index.
[b]A field technician rated home condition at the time of the home visit using a five-point scale that we dichotomized into higher vs. lower.
[c]Current defined as past 12 months.
[d]No category represents the remaining three seasons combined except for the season of interest. For example, samples collected during summer, fall, or winter were assigned to No group for the analysis of spring; samples collected during spring, fall, or winter were assigned to No group for the analysis of summer.
[e]p-Value from linear regression model examining associations of an exposure (predictor) with the bacterial diversity (Shannon index; response variable).
[f]p-Value for linear trend from regression using the number of types of animals (ordered) as the predictor and the Shannon index as the response variable.

Many OTUs were differentially abundant in relation to environmental exposures (Table 3). Mean abundances differed significantly in 94 OTUs for the presence of an indoor dog (vs. no indoor dog), 45 OTUs for home condition (higher vs. lower), and 23 OTUs for winter (vs. other seasons combined) (Table S5). Notably, many more OTUs showed differential abundance for current farming: 388 for crop farming vs. no crop farming and 425 for animal farming vs. no animal farming These amounted to 534 unique OTUs: 109 related to crop farming only, 146 to animal farming only, and 279 to both crop and animal farming (Table S6). When we analyzed the four-level combined farming variable (based on crop and animal farming), we found significant differential abundance for 631 OTUs (Table S7) after adjusting for presence of an indoor dog, home condition, and winter. Summary of number of OTUs differentially abundant by the exposures at the phylum level can be found in Table 3 and Figure 2. Among these OTUs, the top phylum, based on the number of significant OTUs within the phylum, was Proteobacteria, followed by Firmicutes and Actinobacteria. When we assessed differences in mean relative abundance of the families to which the 631 OTUs belong according to the four categories of current farming (no farming, crop farming only, animal farming only, both crop and animal farming), the top five families were Moraxellaceae, Clostridiaceae, Prevotellaceae, Propionibacteriaceae, and Bacillaceae (Table S8). Of interest, there were two phyla, Chloroflexi and Verrucomicrobia, in which OTUs were not related to other factors (presence of an indoor dog, home condition, or winter), but were significantly associated with current farming (Table 3). The phylum Cyanobacteria contained OTUs associated with home condition, winter, and crop and animal farming, but not related to the presence of a dog. Some OTUs in Fusobacteria were associated with the presence of an indoor dog and home condition, but not related to either crop or animal farming.

Some OTUs that were differentially abundant for presence of an indoor dog, home condition, winter, crop farming, or animal farming were unique to each exposure: 71 unique to indoor dog exposure, 10 unique to home condition, and 11 unique to winter vs. other seasons combined (Table S9).

Many OTUs were uniquely related to either crop farming or animal farming but not related to other exposures (Table S10). Unique to crop farming were 103 OTUs: 101 assigned to 49 families within eight phyla, and two unassigned. The top two phyla uniquely related to crop farming were Proteobacteria (36%) and Actinobacteria (22%). Of the 49 families, the top five families based on the number of significantly differentially abundant OTUs within the family were Sphingomonadaceae, Xenococcaceae, Micromonosporaceae, Sphingobacteriaceae, and Nostocaceae. All five families were less abundant in samples from homes of participants doing crop farming than in those with no crop farming exposure (Table 4). There were 133 OTUs unique to animal farming, including 130 assigned to 71 families within 7 phyla, and 3 unassigned (Table S10). For animal farming, the top two phyla uniquely related to this exposure were Firmicutes (28%) and Proteobacteria (22%). Of the 71 families uniquely related to animal farming, the top five families were Bacillaceae, Bacteroidaceae, Xanthomonadaceae, Streptococcaceae, and Lactobacillaceae (Table 4). Of the five, three (Bacillaceae, Xanthomonadaceae, and Lactobacillaceae) were more abundant in house dust from individuals working with farm animals than from those not working with farm animals; the other two were less abundant in samples from those doing animal farming. There were 254 OTUs differentially abundant for crop farming (vs. no

Table 3. Number of operational taxonomic units (OTUs) in each phylum showing significant differential abundance [false discovery rate (FDR) < 0.05] for each of several exposures.

	Exposures									
Phylum	Presence of an indoor dog (vs. no indoor dog)	Home condition (higher vs. lower)[a]	Winter (vs. other seasons combined)	Crop farming (vs. no crop farming)	Crop farming (vs. no crop farming)[b]	Animal farming (vs. no animal farming)	Animal farming (vs. no animal farming)[c]	Both crop and animal farming,[d] unadjusted	Both crop and animal farming, adjusted[e]	Total
Proteobacteria	24	14	8	98	37	91	31	163	159	376
Firmicutes	32	15	6	77	4	116	37	152	146	315
Actinobacteria	7	5	0	73	23	76	26	123	122	239
Bacteroidetes	23	6	2	77	13	90	25	112	114	237
Cyanobacteria	0	3	4	30	18	12	0	31	30	52
Chloroflexi	0	0	0	14	0	18	4	21	21	34
Acidobacteria	0	0	1	5	4	6	5	11	11	28
Fusobacteria	5	2	0	0	0	0	0	1	0	14
Verrucomicrobia	0	0	0	2	1	3	2	5	4	13
[Thermi]	1	0	0	0	0	1	0	2	3	12
TM7	1	0	0	3	0	3	0	4	4	11
Euryarchaeota[f]	0	0	0	2	0	2	0	3	3	5
Tenericutes	1	0	0	0	0	0	0	0	0	4
Gemmatimonadetes	0	0	0	0	0	0	0	0	0	4
FBP	0	0	0	1	0	1	0	2	2	3
Armatimonadetes	0	0	0	1	1	0	0	1	1	3
Crenarchaeota[f]	0	0	0	0	0	0	0	1	1	2
Planctomycetes	0	0	0	0	0	0	0	1	1	1
Phylum unassigned	0	0	2	5	2	6	3	9	9	32
Total	94	45	23	388	103	425	133	642	631	1385

Note: OTUs that show significant differential abundance were identified by using analysis of composition of microbiomes (ANCOM).
[a]A field technician rated home condition at the time of the home visit using a five-point scale that we dichotomized into higher vs. lower for analysis.
[b]Uniquely associated with crop farming (vs. no crop farming): no univariate association with other factors: presence of an indoor dog (vs. no indoor dog), home condition (higher vs. lower), winter (vs. other seasons combined), and animal farming (vs. no animal farming).
[c]Uniquely associated with animal farming (vs. no animal farming): no univariate association with other factors: presence of an indoor dog (vs. no indoor dog), home condition (higher vs. lower), winter (vs. other seasons combined), and crop farming (vs. no crop farming).
[d]Four-level current farming variable: neither crop nor animal farming, crop farming only, animal farming only, both crop and animal farming.
[e]Four-level current farming variable: neither crop nor animal farming, crop farming only, animal farming only, both crop and animal farming: after adjusting for presence of an indoor dog (vs. no indoor dog), home condition (higher vs. lower), and winter (vs. other seasons combined).
[f]Euryarchaeota and Crenarchaeota are in the kingdom Archaea. The rest of the phyla are in the kingdom Bacteria.

crop farming) as well as animal farming (vs. no animal farming) that were not related to the other nonfarm exposures, including presence of an indoor dog, home condition, and winter vs. other seasons combined (Table S11).

Endotoxin levels were higher in Iowa than North Carolina and in homes rated in the lower home condition category. Endotoxin concentrations were higher in homes with indoor pets or occupants doing crop or animal farming than in homes without these exposures (Table S12). After adjusting for the nonfarm factors significantly associated with endotoxin in our data (presence of indoor pets and home condition), the positive associations between endotoxin and current farming remained significant (p-value = 0.005 for crop farming; p-value = 9.5×10^{-5} for animal farming) (Table S13). There was a trend for increasing Shannon index (alpha diversity) across quartiles of endotoxin (p-value for linear trend = 0.009) (Table S14). We found 257 OTUs within 10 phyla that increased across endotoxin quartiles and 452 OTUs within 17 phyla that decreased across endotoxin categories (Table S15 and Table S16). Of note, 82 of the 257 (32%) positively associated OTUs, and 105 (23%) of the 452 negatively associated OTUs were from Proteobacteria, a major group of Gram-negative bacteria (Figure S1). None of the positively associated OTUs were from Cyanobacteria, which stain Gram-negative but differ from fecal related Gram-negative bacteria in the lipopolysaccharide in their cell wall (Stewart et al. 2006).

One of potential factors reported to be associated with endotoxin is carpeting (Douwes et al. 2000; Mendy et al. 2018). In our study, nearly all of our samples were collected from floors with carpet: 94% overall, with little difference between Iowa (95%) and North Carolina (91%). The mean Shannon index was slightly lower in samples from carpeted surface (4.57, SD = 0.76) compared with samples from smooth floor (4.73, SD = 0.65), but the difference was not significant ($p = 0.13$). Although there was no significant association between diversity and carpeted surface vs. smooth floor, we tested whether associations differed when adding a term for carpet to the adjusted models and they did not: for living on a farm, $p = 0.525$ before and $p = 0.518$ after adjustment; for crop farming, $p = 0.006$ before and 0.006 after adjustment; and for animal farming, $p = 3.4 \times 10^{-5}$ before and $p = 4.8 \times 10^{-5}$ after adjustment. Endotoxin levels were higher in dust samples collected from carpet [Geometric Mean (GM) = 39.2; Geometric Standard Deviation (GSD) = 3.8] vs. smooth floor (GM = 30.0; GSD = 7.1), but the difference was not statistically significant ($p = 0.15$). Given the nearly universal presence of carpet in the room from which we collected the dust sample, we were underpowered to identify associations with carpeting compared to the larger NHANES (National Health and Nutrition Examination Survey) study of endotoxin levels ($n = 6{,}963$).

Discussion

To our knowledge, this is the first large-scale study to examine associations between the house dust microbiota and both farm and nonfarm exposures using *16S rRNA* gene amplicon sequencing in an agricultural population. Given the substantial body of literature on both the environmental determinants and health effects of house dust endotoxin, we also examined associations between house dust endotoxin and house dust microbiota. Our findings suggest that the indoor microbiota differs between dust samples from homes of individuals with current farming exposure and those from individuals who are not currently farming, with additional variation depending on specific farming types. Specifically, crop and animal farming were associated

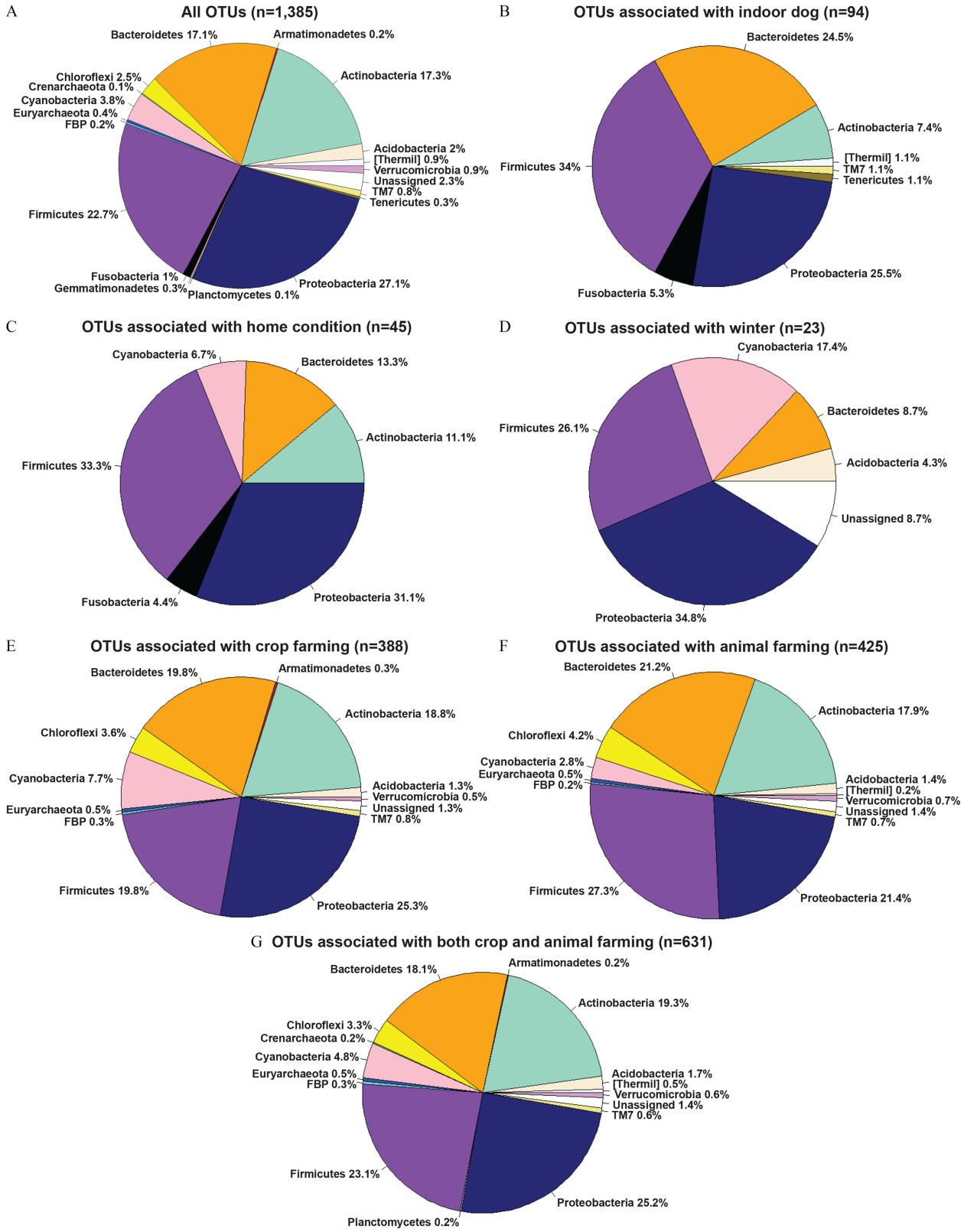

Figure 2. Proportion of operational taxonomic units (OTUs) significantly associated [false discovery rate (FDR) < 0.05] with each exposure at the phylum level. Pie chart shows proportions of significant OTUs assigned to each phylum in relation to each exposure. Phyla having more than one OTUs were shown in the pie chart. Of all OTUs ($n = 1,385$) in our data, there were (*A*) 1,353 assigned to 18 phyla, 32 OTUs unassigned. Of the 1,385 OTUs, (*B*) 94 within 8 phyla were significantly associated with presence of an indoor dog (vs. no indoor dog); (*C*) 45 within 6 phyla were associated with home condition (higher vs. lower); (*D*) 21 within 5 phyla and 2 unassigned were associated with winter (vs. other seasons combined); (*E*) 383 within 12 phyla and 5 unassigned were associated with crop farming (vs. no crop farming); (*F*) 419 within 12 phyla and 6 unassigned were associated with animal farming (vs. no animal farming); and (*G*) 622 within 15 phyla and 9 unassigned were associated with both crop and animal farming (four-level combined variable) adjusted for presence of an indoor dog (vs. no indoor dog), home condition (higher vs. lower), and winter (vs. other seasons combined).

Table 4. Relative abundance of families including at least two significant operational taxonomic units (OTUs) [false discovery rate (FDR) < 0.05] uniquely associated with either crop or animal farming.

		Relative abundance			No. of OTUs in family		
		Exposure					
Phylum	Family	Yes (%)	No (%)	Direction[a]	Total	No. significant	No. greater relative abundance in the exposed group[b]
Exposure: crop farming (Yes) vs. no crop farming (No)							
Proteobacteria	Sphingomonadaceae	0.120	0.202	-	44	9	0
Cyanobacteria	Xenococcaceae	0.035	0.086	-	13	6	0
Actinobacteria	Micromonosporaceae	0.174	0.252	-	11	6	0
Bacteroidetes	Sphingobacteriaceae	0.109	0.155	-	25	6	0
Cyanobacteria	Nostocaceae	0.033	0.088	-	7	5	0
Bacteroidetes	Cytophagaceae	0.121	0.092	+	39	4	2
Proteobacteria	Acetobacteraceae	0.024	0.047	-	20	4	0
Actinobacteria	Actinomycetales (O)[c]	0.024	0.047	-	18	4	0
Proteobacteria	Caulobacteraceae	0.048	0.077	-	14	4	0
Proteobacteria	Methylocystaceae	0.031	0.056	-	8	4	0
Bacteroidetes	Chitinophagaceae	0.055	0.076	-	24	3	3
Proteobacteria	Burkholderiaceae	0.027	0.049	-	6	3	1
Acidobacteria	Acidobacteriaceae	0.012	0.031	-	4	3	0
Firmicutes	Bacillaceae	0.222	0.387	-	27	2	0
Actinobacteria	Gaiellaceae	0.041	0.038	+	12	2	1
Proteobacteria	Oxalobacteraceae	0.099	0.141	-	9	2	0
Cyanobacteria	Chlorophyta (O)[c]	0.049	0.072	-	7	2	0
Actinobacteria	Solirubrobacteraceae	0.034	0.045	-	4	2	0
Phylum unassigned	Family unassigned	0.039	0.032	+	32	2	1
Exposure: animal farming (Yes) vs. no animal farming (No)							
Firmicutes	Bacillaceae	0.168	0.116	+	27	8	8
Bacteroidetes	Bacteroidaceae	0.811	1.208	-	25	7	2
Proteobacteria	Xanthomonadaceae	0.057	0.040	+	26	5	5
Firmicutes	Lactobacillaceae	0.707	0.682	+	13	4	4
Firmicutes	Streptococcaceae	0.709	0.940	-	11	4	1
Actinobacteria	Acidimicrobiales (O)[c]	0.057	0.041	+	7	4	4
Firmicutes	Ruminococcaceae	0.046	0.014	+	46	3	3
Bacteroidetes	Flavobacteriaceae	0.207	0.174	+	33	3	3
Bacteroidetes	Prevotellaceae	0.041	0.015	+	27	3	3
Actinobacteria	Nocardioidaceae	0.316	0.247	+	25	3	3
Bacteroidetes	Chitinophagaceae	0.033	0.028	+	24	3	3
Firmicutes	Aerococcaceae	0.209	0.128	+	16	3	3
Proteobacteria	Comamonadaceae	0.087	0.060	+	12	3	3
Firmicutes	Planococcaceae	0.063	0.041	+	8	3	3
Proteobacteria	Erythrobacteraceae	0.064	0.045	+	7	3	3
Actinobacteria	Streptomycetaceae	0.074	0.047	+	4	3	3
Firmicutes	Lachnospiraceae	0.016	0.009	+	60	2	2
Firmicutes	[Tissierellaceae]	0.326	0.467	-	27	2	0
Firmicutes	Clostridiales (O)[c]	0.028	0.012	+	25	2	2
Bacteroidetes	Sphingobacteriaceae	0.055	0.040	+	25	2	2
Proteobacteria	Acetobacteraceae	0.072	0.119	-	20	2	0
Proteobacteria	Caulobacteraceae	0.189	0.152	+	14	2	2
Chloroflexi	JG30-KF-CM45 (O)[c]	0.039	0.029	+	13	2	2
Proteobacteria	Alcaligenaceae	0.061	0.044	+	9	2	2
Acidobacteria	iii1–15 (O)[c]	0.058	0.048	+	8	2	2
Acidobacteria	Ellin6075	0.041	0.031	+	8	2	2
Actinobacteria	Intrasporangiaceae	0.117	0.068	+	7	2	2
Chloroflexi	Ellin6529 (C)[c]	0.037	0.031	+	5	2	2
Actinobacteria	Cellulomonadaceae	0.128	0.102	+	3	2	2
Actinobacteria	Promicromonosporaceae	0.041	0.022	+	2	2	2
Phylum unassigned	Family unassigned	0.243	0.180	+	32	3	3

Note: Relative abundance $(\%)$ of families (in a sample) $= \frac{\text{sequence counts of significant OTUs in the family}}{\text{total sequence counts}} * 100$ The OTUs were not related to other factors, including presence of an indoor dog, home condition, and winter (vs. other seasons combined).
[a]Direction indicates increased (+) or decreased (-) relative abundance in a group exposed to a farm factor crop or animal compared to a group unexposed to the farm factor (NO).
[b]Higher relative abundances in the farming group (Yes) than the no farming group (No). For example, higher relative abundance in crop farming group vs. no crop farming group.
[c]For unassigned family, we noted assigned class with (C). When both family and class were unassigned, we noted assigned order with (O).

with higher bacterial diversity. We identified specific OTUs that were higher or lower in abundance in the homes of participants who are currently farming than in homes of unexposed participants. We identified associations between specific OTUs and endotoxin levels, which had the expected positive relationship with current farming.

Most studies exploring determinants underlying the indoor microbiota have been conducted in nonfarming populations. These studies have reported that occupant-related factors such as the ratio of females to males in households (Barberán et al. 2015) and the presence of indoor pets (Barberán et al. 2015; Dannemiller et al. 2016) have greater influence on the indoor bacterial communities than outdoor sources such as geographic (Adams et al. 2014; Barberán et al. 2015) or seasonal factors (Adams et al. 2014; Weikl et al. 2016). We confirmed the previously reported associations between indoor bacterial communities and indoor sources, including presence of an indoor dog (Barberán et al. 2015; Dannemiller et al. 2016) and home

condition (Dunn et al. 2013); we had no data on gender ratio. In addition, bacterial communities in dust samples collected from the bedrooms of participants who were currently farming differed from those in dust samples collected from the bedrooms of participants not currently farming. In a study of 86 children's homes, Birzele and coauthors reported that living on a farm can affect the house dust microbiota (Birzele et al. 2017). Compared to that study where only 13% lived on a farm, most of our participants had some farm exposure. Nearly everyone ($n = 844$; 96%) reported ever having done farm work. The remaining 35 who reported no prior farm work were all female, including 34 who are spouses of farmers. The vast majority of individuals (83%) reported currently living on a farm. The other 152 who reported not currently living on a farm include 54 who nonetheless reported current farming. Given the older age of many of our participants, it is perhaps not surprising that a small proportion are no longer living on the farm. Stein and coauthors reported that indoor bacterial communities differed between two populations practicing either traditional or industrialized farming based on analysis of one pooled sample from each group (Stein et al. 2016). Although it is possible that these bacteria enter the home via airborne spread from the adjacent farm, they may also be carried into the home on the skin or clothing of the farming occupants.

At the phylum level, the top two phyla associated with current farming in our study have been implicated in human health: Proteobacteria and Firmicutes. Proteobacteria, a phylum of Gram-negative bacteria, includes bacteria found in the normal human microbiota as well as pathogens such as *Salmonella*, *Vibrio*, and *Helicobacter*. Firmicutes, which are Gram-positive, comprise the largest portion of human gut microbiota (Ley et al. 2006) and have been associated with obesity and other health outcomes (Turnbaugh et al. 2006).

Many earlier house dust studies (Liebers et al. 2008) measured endotoxin concentration as a proxy for bacterial exposures, but how endotoxin levels relate to the house dust microbiota is unknown. We confirmed previously reported associations of house dust endotoxin levels with presence of indoor pets, better home condition, geographical region, and farm exposures (Chen et al. 2012; Giovannangelo et al. 2007; Holst et al. 2015; Waser et al. 2004), but also examined how endotoxin relates to the microbiota. Many OTUs from Proteobacteria were positively associated with the endotoxin, which can be expected, given that Proteobacteria is a major group of Gram-negative bacteria.

We recognize the possibility that the farmer's spouses (99.5% female) in our study who reported that they were not farming may have had exposure to farming from their spouse's work. This would tend to attenuate associations between bacterial diversity and farming. Because our survey did not include questions regarding current farming of participants' spouses, we cannot address this question directly.

When examining associations of bacterial communities with farm exposures, we did not adjust for state of residence (North Carolina vs. Iowa) because there were substantial differences in current farming exposures by state. Of the 596 participants in Iowa, 85% reported living on a farm, 65% reported crop farming, 55% reported animal farming, and 72% reported either crop or animal farming (Table S1). Of the 283 participants in North Carolina, 78% reported living on a farm, 33% reported crop farming, 40% reported animal farming, and 51% reported either crop or animal farming. Thus, instead of adjusting for state, we performed a stratified analysis; we found similar associations between exposures and bacterial communities in Iowa and North Carolina (Table S1).

Our study has other limitations. We have only a single sample of house dust. Thus, we assume the samples reflect the usual home condition. We are unable to quantify absolute levels of bacteria in dust; however, we were able to quantify relative abundances of bacterial taxa. Our rating of the home condition was based on a five-level scale using an instrument from a previous NIEHS study (Arbes et al. 2003). This rating is subjective, and we did not assess interrater reliability.

A strength of this study is the large size compared to sizes, ranging between 86 and 196, of most earlier studies (Birzele et al. 2017; Dannemiller et al. 2016) of indoor microbiota using high-throughput sequencing methods, except for one larger study (Barberán et al. 2015) with 1,142 indoor dust samples that examined only nonfarm exposures. An additional strength is our ability to examine the contribution of both farm and nonfarm exposures. We could assess both farm exposures, with detailed information on two types of farming (crop and/or animal farming) and four types of farm animals, and nonfarm exposures that have been related to indoor microbiota, including the presence of dogs or cats inside the home and home condition assessed by an objective observer. We also included two outdoor factors: state of residence and season of dust collection. Endotoxin exposure assessment was conducted with high levels of quality assurance. Although we measured only bedroom dust, most individuals spend a large portion of their day in the bedroom, making this a highly relevant single location to sample.

Many of our results are reassuringly consistent with what we predicted. For example, more OTUs from Cyanobacteria, a phylum of photosynthetic bacteria that exist in moist soil, were associated with crop farming than with animal farming. Two phyla containing OTUs associated with current farming, but not with other factors, were Chloroflexi, abundant in agricultural soil (Chapagain and Good 2015), and Verrucomicrobia, mostly found in freshwater and soil. Moreover, two OTUs assigned to the genus *Fusobacterium* were associated only with presence of an indoor dog and home condition, and not with farming or other factors. Bacteria from this genus are found in most dogs and cats; they can cause severe infections after a dog bite. Of the bacterial families showing elevated relative abundances in relation to animal farming, some are related to animals in the literature. For example, the family Bacillaceae includes the species *Bacillus anthracis* that causes anthrax, a disease associated with animal exposures (Spencer 2003). Lactobacillaceae is found in dairy and grain products, water, and soil.

Conclusions

In this agricultural population, we found that current farming, including both crop and animal farming, presence of an indoor dog, and home condition, were significant predictors of the composition of the house dust microbiota. Current farming was associated with the microbiota inside homes, even after adjusting for indoor environmental factors, including pets and home condition. Many taxa related to current farming were also related to house dust endotoxin, a widely studied surrogate of bacterial exposure that has been related to various health outcomes. These results suggest that indoor microbial signatures might serve as markers for unique and specific exposures to microbes associated with crops or farm animals that cannot be obtained by simply asking individuals about these farm exposures. This comprehensive investigation of factors that predict the bacterial communities inside homes, including current crop and animal farming, and predictors common to both farming and nonfarming populations, is an essential step toward understanding the impact of exposure to the indoor dust microbiota on human health.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health (NIH), the National Institute of Environmental Health Sciences (NIEHS) (Z01-ES049030 and

Z01-ES102385), the National Cancer Institute (Z01-CP010119B), and by American Recovery and Reinvestment Act funds. The Microbiome Core Facility at the University of North Carolina is supported in part by NIH National Institute of Diabetes and Digestive and Kidney Diseases grant P30 DK34987. P.S.T. was supported by PHR-SUPS2-S-10-00179 and NIH P30 ES005605. We appreciate all of the study participants for their contribution to this research. We thank Drs. F. Day of NIEHS for expert computational assistance and J. Hoppin (North Carolina State University, Raleigh, NC) for her important contribution to the Agricultural Lung Health Study during her tenure at NIEHS.

References

Adams RI, Miletto M, Lindow SE, Taylor JW, Bruns TD. 2014. Airborne bacterial communities in residences: similarities and differences with fungi. PLoS One 9(3):e91283, PMID: 24603548, https://doi.org/10.1371/journal.pone.0091283.

Alavanja MC, Sandler DP, McMaster SB, Zahm SH, McDonnell CJ, Lynch CF, et al. 1996. The agricultural health study. Environ Health Perspect 104(4):362–369, PMID: 8732939, https://doi.org/10.1289/ehp.96104362.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol 26(1):32–46, https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.

Arbes SJ Jr., Sever M, Archer J, Long EH, Gore JC, Schal C, et al. 2003. Abatement of cockroach allergen (BLA g 1) in low-income, urban housing: a randomized controlled trial. J Allergy Clin Immunol 112(2):339–345, PMID: 12897740, https://doi.org/10.1067/mai.2003.1597.

Barberán A, Dunn RR, Reich BJ, Pacifici K, Laber EB, Menninger HL, et al. 2015. The ecology of microscopic life in household dust. Proc R Soc 282(1814):1139–1220, PMID: 26311665, https://doi.org/10.1098/rspb.2015.1139.

Birzele LT, Depner M, Ege MJ, Engel M, Kublik S, Bernau C, et al. 2017. Environmental and mucosal microbiota and their role in childhood asthma. Allergy 72(1):109–119, PMID: 27503830, https://doi.org/10.1111/all.13002.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 10(1):57–59, PMID: 23202435, https://doi.org/10.1038/nmeth.2276.

Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27(4):325–349, https://doi.org/10.2307/1942268.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010a. PyNAST: A flexible tool for aligning sequences to a template alignment. Bioinformatics 26(2):266–267, PMID: 19914921, https://doi.org/10.1093/bioinformatics/btp636.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. 2010b. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336, PMID: 20383131, https://doi.org/10.1038/nmeth.f.303.

Carnes MU, Hoppin JA, Metwali N, Wyss AB, Hankinson JL, O'Connell EL, et al. 2017. House dust endotoxin levels are associated with adult asthma in a U.S. farming population. Annals ATS 14(3):324–331, PMID: 27977294, https://doi.org/10.1513/AnnalsATS.201611-861OC.

Chapagain T, Good A. 2015. Yield and production gaps in rainfed wheat, barley, and canola in Alberta. Front Plant Sci 6:990, PMID: 26635824, https://doi.org/10.3389/fpls.2015.00990.

Chen CM, Thiering E, Doekes G, Zock JP, Bakolis I, Norbäck D, et al. 2012. Geographical variation and the determinants of domestic endotoxin levels in mattress dust in Europe. Indoor Air 22(1):24–32, PMID: 21906176, https://doi.org/10.1111/j.1600-0668.2011.00740.x.

Clarke KR. 1993. Non-parametric multivariate analysis of changes in community structure. Austral Ecol 18(1):117–143, https://doi.org/10.1111/j.1442-9993.1993.tb00438.x.

Dannemiller KC, Gent JF, Leaderer BP, Peccia J. 2016. Influence of housing characteristics on bacterial and fungal communities in homes of asthmatic children. Indoor Air 26(2):179–192, PMID: 25833176, https://doi.org/10.1111/ina.12205.

Dassonville C, Demattei C, Vacquier B, Bex-Capelle V, Seta N, Momas I. 2008. Indoor airborne endotoxin assessment in homes of Paris newborn babies. Indoor Air 18(6):480–487, PMID: 19120498, https://doi.org/10.1111/j.1600-0668.2008.00549.x.

Douwes J, Zuidhof A, Doekes G, van der Zee SC, Wouters I, Boezen MH, et al. 2000. (1–>3)-beta-d-glucan and endotoxin in house dust and peak flow variability in children. Am J Respir Crit Care Med 162(4 Pt 1):1348–1354, PMID: 11029343, https://doi.org/10.1164/ajrccm.162.4.9909118.

Dunn RR, Fierer N, Henley JB, Leff JW, Menninger HL. 2013. Home life: factors structuring the bacterial diversity found within and between homes. PLoS One 8(5):e64133, PMID: 23717552, https://doi.org/10.1371/journal.pone.0064133.

Edgar RC. 2010. Search and clustering orders of magnitude faster than blast. Bioinformatics 26(19):2460–2461, PMID: 20709691, https://doi.org/10.1093/bioinformatics/btq461.

Giovannangelo M, Gehring U, Nordling E, Oldenwening M, Terpstra G, Bellander T, et al. 2007. Determinants of house dust endotoxin in three European countries - the AIRALLERG study. Indoor Air 17(1):70–79, PMID: 17257154, https://doi.org/10.1111/j.1600-0668.2006.00461.x.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. 2011. Chimeric 16s RRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res 21(3):494–504, PMID: 21212162, https://doi.org/10.1101/gr.112730.110.

Holst G, Høst A, Doekes G, Meyer HW, Madsen AM, Sigsgaard T. 2015. Determinants of house dust, endotoxin, and beta-(1→3)-d-glucan in homes of Danish children. Indoor Air 25(3):245–259, PMID: 25039673, https://doi.org/10.1111/ina.12143.

House JS, Wyss AB, Hoppin JA, Richards M, Long S, Umbach DM, et al. 2017. Early-life farm exposures and adult asthma and atopy in the Agricultural Lung Health Study. J Allergy Clin Immunol 140(1):249–256.e14, PMID: 27845237, https://doi.org/10.1016/j.jaci.2016.09.036.

Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124(4):837–848, PMID: 16497592, https://doi.org/10.1016/j.cell.2006.02.017.

Liebers V, Raulf-Heimsoth M, Brüning T. 2008. Health effects due to endotoxin inhalation (review). Arch Toxicol 82(4):203–210, PMID: 18322674, https://doi.org/10.1007/s00204-008-0290-1.

Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71(12):8228–8235, PMID: 16332807, https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 26:27663, PMID: 26028277, https://doi.org/10.3402/mehd.v26.27663.

McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8(4):e61217, PMID: 23630581, https://doi.org/10.1371/journal.pone.0061217.

Mendy A, Wilkerson J, Salo PM, Cohn RD, Zeldin DC, Thorne PS. 2018. Endotoxin predictors and associated respiratory outcomes differ with climate regions in the U.S. Environ Int 112:218–226, PMID: 29277065, https://doi.org/10.1016/j.envint.2017.12.003.

Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, et al. 2013. Advancing our understanding of the human microbiome using QIIME. Meth Enzymol 531:371–444, PMID: 24060131, https://doi.org/10.1016/B978-0-12-407863-5.00019-8.

Ownby DR, Johnson CC, Peterson EL. 2002. Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age. JAMA 288(8):963–972, PMID: 12190366, https://doi.org/10.1001/jama.288.8.963.

Peddada S, Harris S, Zajd J, Harvey E. 2005. ORIOGEN: order restricted inference for ordered gene expression data. Bioinformatics 21(20):3933–3934, PMID: 16109745, https://doi.org/10.1093/bioinformatics/bti637.

Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26(7):1641–1650, PMID: 19377059, https://doi.org/10.1093/molbev/msp077.

Shannon CE. 1948. A mathematical theory of communication. At&T Tech J 27(3):379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Spencer RC. 2003. Bacillus anthracis. J Clin Pathol 56(3):182–187, PMID: 12610093, https://doi.org/10.1136/jcp.56.3.182.

Stein MM, Hrusch CL, Gozdz J, Igartua C, Pivniouk V, Murray SE, et al. 2016. Innate immunity and asthma risk in Amish and Hutterite farm children. N Engl J Med 375(5):411–421, PMID: 27518660, https://doi.org/10.1056/NEJMoa1508749.

Stewart I, Schluter PJ, Shaw GR. 2006. Cyanobacterial lipopolysaccharides and human health - a review. Environ Health 5:7, PMID: 16563160, https://doi.org/10.1186/1476-069X-5-7.

Thorne PS, Kulhánková K, Yin M, Cohn R, Arbes SJ Jr., Zeldin DC. 2005. Endotoxin exposure is a risk factor for asthma: the national survey of endotoxin in United States housing. Am J Respir Crit Care Med 172(11):1371–1377, PMID: 16141442, https://doi.org/10.1164/rccm.200505-758OC.

Thorne PS, Mendy A, Metwali N, Salo P, Co C, Jaramillo R, et al. 2015. Endotoxin exposure: Predictors and prevalence of associated asthma outcomes in the united states. Am J Respir Crit Care Med 192(11):1287–1297, PMID: 26258643, https://doi.org/10.1164/rccm.201502-0251OC.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444(7122):1027–1031, PMID: 17183312, https://doi.org/10.1038/nature05414.

Vojta PJ, Friedman W, Marker DA, Clickner R, Rogers JW, Viet SM, et al. 2002. First national survey of lead and allergens in housing: survey design and

methods for the allergen and endotoxin components. Environ Health Perspect 110(5):527–532, PMID: 12003758, https://doi.org/10.1289/ehp.02110527.

von Mutius E, Vercelli D. 2010. Farm living: effects on childhood asthma and allergy. Nat Rev Immunol 10(12):861–868, PMID: 21060319, https://doi.org/10.1038/nri2871.

Waser M, Schierl R, Von Mutius E, Maisch S, Carr D, Riedler J, et al. 2004. Determinants of endotoxin levels in living environments of farmers' children and their peers from rural areas. Clin Exp Allergy 34(3):389–397, PMID: 15005732, https://doi.org/10.1111/j.1365-2222.2004.01873.x.

Weikl F, Tischer C, Probst AJ, Heinrich J, Markevych I, Jochner S, et al. 2016. Fungal and bacterial communities in indoor dust follow different environmental determinants. PLoS One 11(4):e0154131, PMID: 27100967, https://doi.org/10.1371/journal.pone.0154131.